Standardized Workflows for Near Complete Proteome Identification

Xiaoyue Jiang[1], Shouling Xu[2,3], Zhiyong Wang[2] , Al Burlingame[3] ,Andreas FR Hühmer[1]

[1]Thermo Fisher Scientific, San Jose, CA [2]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA [3]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA

Overview

Purpose: Standardized workflow solutions for near complete identification of proteomes are developed. The performance of the standardized workflows for various samples is evaluated.

Methods: The standardized workflows composed of sample prep, LC-MS and data processing methods are evaluated using UPS2, HeLa digest and phosphorylated peptide samples.

Results: The standardized solution can provide comparable performance to optimal or published results.

Introduction

Efforts during the past 15 years to characterize protein dynamics using proteomics have led to a better understanding of molecular drivers for complex biological processes and diseases. However, our ability to translate proteomic results to clinical success has been remarkably low. One big reason is the quality of the proteomics data, mostly due to the technical difficulties during the validation phase. This is a consequence of the steep learning curve of the technology, the lack of standardized methods, and the complexity of the experiments.

By developing a standardized workflow solution based on Q Exactive technology, we aim to overcome these challenges, and facilitate access of the technology to non-expert users. Our result shows that standard methods compare favorably with results analyzed with optimized methods and we can achieve near complete proteome identification with standard methods.

Methods

The workflow solution standardizes wet chemistry kits and protocols, LC-MS hardware components, instrument methods and processing workflows. An overview of the solution components is illustrated in Figure 1.

Sample Preparation

Peptide Retention Time Calibration Mixture (PRTC, Thermo Fisher, IL) and HeLa Protein Digest Standard (Thermo Fisher, IL) are used to monitor the instrument status. Several injections are needed to assure the reproducibility, separation efficiency and instrument sensitivity. We also standardize the protein digestion protocol by using Pierce Mass Spec Sample Prep Kit for Cultured Cells (Thermo, IL) and fractionation protocol using High pH Reversed-Phase Fractionation Kit prototype (Thermo, IL).

Liquid Chromatography

A Thermo Scientific™ EASY-nLC™ 1000 UPLC system and Thermo Scientific™ EASY-Spray™ Source with 50 cm EASY-Spray Column is used to separate peptides. The gradient is fixed as 1hour, 2hour and 3hour lengths for different types of workflow. The temperature of the column is set as 40C.

Mass Spectrometry

The samples are analyzed using higher-energy C-trap dissociation (HCD) data-dependent tandem MS method on a Thermo Scientific™ Q Exactive™ Plus mass spectrometer. Instrument methods of peptide identification are modified based on fast and sensitive scanning methods in the literature [1]. The sample load and complexity are considered when choosing the standard methods.

Data Analysis

The UPS2 and HeLa data are analyzed by Thermo Scientific™ Proteome Discoverer™ software (v.2.0.) with the SEQUEST HT search engine. The phosphoprotein data is searched by ProteinProspector v5.14.0 (not part of the standard solution).The workflow and modification details can be found in [2]. Resulting peptide hits are filtered for maximum 1% FDR.

FIGURE 1. An overview of the standardized workflow solution.

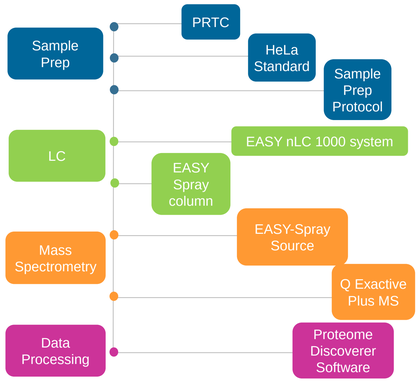


Table 1. Supported proteomics identification applications by standardized workflow solution.

	Workflows
Quality Control	PRTC Quality Control
	HeLa Digest Quality Control
	Blank/Wash
Bottom-up Proteomics Identification	General Peptide Identification
	Whole Proteome Mapping
	Phosphorylated Peptide Identification

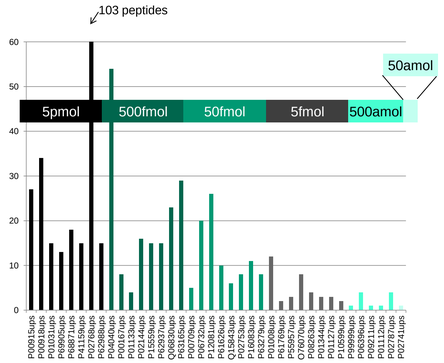
Results

Standardized workflow solution can be used in multiple, widely-adopted LC-MS applications, as demonstrated in Table 1. The results for different workflows are illustrate below.

UPS2

Sigma-Aldrich UPS-2 protein standard mixture contains 48 proteins, spanning a dynamic range of 6 orders of magnitude. We test the dynamic range of our workflow on this challenging sample. The sample is digested with the standard digestion kit, followed by "General Peptide Identification" workflow which is composed of 1hour LC gradient and sensitive scanning MS methods. As shown in Figure 2, we identify 38 proteins covering a concentration of 6 decades. This provides confidence to identify all proteins covered in a biological pathway, which normally spans 3-4 orders of magnitude.

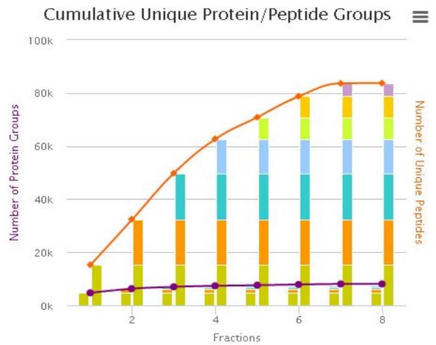
FIGURE 2. Protein identifications in the analysis of UPS2 standard. Single injection covers proteins with six concentration decades. The absolute injection amount is shown on the top.



Whole Proteome Mapping

"Whole Proteome Mapping" workflow is tested on HeLa cell lysate. The same digestion kit as the previous experiment is used. In addition, the digest is fractionated using high PH reverse phase spin columns into 8 fractions. The 2 hour LC gradient and fast scanning MS method is used for each fraction. It takes approximately one day to complete the proteome mapping. We detect 83,751 unique peptides and 8,050 protein groups, as demonstrated by the cumulative plot in Figure 3. The identification number from our standardized workflow solution is comparable with the results published in literature [3], which involves a UPLC system for the first dimension separation.

FIGURE 3. Numbers of protein groups and peptides identified from 10ug of HeLa cell lysate. Colored bars indicate increase in unique identifications per fraction for unique proteins (purple) and unique peptides (orange).



Phosphorylated Peptides

We also evaluate the proteome identification with post translational modifications (PTM) signatures like phosphorylation. Studies on extensive analysis and mapping of phosphorylation sites in Arabidopsis would provide insights and foundations for basic research studies, such as signaling cascades and dynamic regulations, as well as providing databases for cross talk analysis. Enriched and fractionated Arabidopsis flower phosphopetides are divided into two groups to map the PTM occurrence and site location. One group is run in university core facility, using regular method. The other group is run by "phosphorylated peptide identification" standard workflow solution. As shown in Figure 4 and Figure 5, compared to the group 1, the standardized workflow solution on Q Exactive Plus provide >3 times more phosphopeptide and >2 times more phosphoprotein identifications. We have detected 8187 phosphoproteins and 46215 phosphopeptides from Arabidopsis flower tissues in group 2, which constitute the biggest phosphopeptide library in plant so far.

FIGURE 4. Numbers of phosphoprotein groups identified using different methods

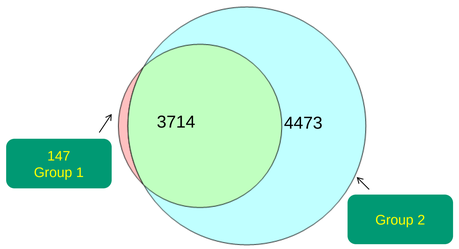
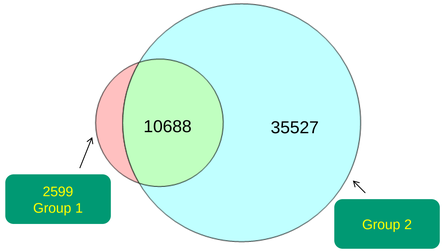


FIGURE 5. Numbers of phosphopeptides identified using different methods



Conclusion

- Standardized proteomics workflow for protein identification are successfully developed
- The performance of the workflow solution for proteomics applications is comparable or better than published results
- Standardized workflow solutions will enable non-LC/MS experts to access state-of-the-art proteomics technologies more readily

References

1. Kelstrup, C. *et al.*, Optimized Fast and Sensitive Acquisition Methods for Shotgun Proteomics on a Quadrupole Orbitrap Mass Spectrometer. *J. Proteome Res.*, **2012**, *11* (6), pp 3487–3497
2. Jiang, X. *et al.*, Achieving the Highest Performance for Protein Identification and SILAC Relative Quantitation on a Benchtop Quadrupole Ultra-high-field Orbitrap mass spectrometer. *Thermo Fisher Poster*, **2014**, ASMS
3. Ding, C. *et al.*, A Fast Workflow for Identification and Quantification of Proteomes. Molecular & Cellular Proteomics, **2013**, *12*, pp 2370-2380

Acknowledgements

We would like to thank Ryan Bomgarden at Thermo Fisher Scientific for supplying the HeLa cell samples.

List all non-Thermo trademarks and registered trademarks that appear in the poster. Examples include TMT, SEQUEST, ActiveX, Eksigent. Change text to black after you have completed the trademark section., Mascot. All other trademarks are the property of Thermo Fisher Scientific and its subsidiaries

This information is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others.